

Acoustic analysis of voice using WPCVox: a comparative study with Multi Dimensional Voice Program

Juan Ignacio Godino-Llorente · Víctor Osma-Ruiz ·
Nicolás Sáenz-Lechón · Ignacio Cobeta-Marco ·
Ramón González-Herranz · Carlos Ramírez-Calvo

Abstract In this study, two different tools developed for the parametric extraction and acoustic analysis of voice samples are compared. The main goal of the paper is to contrast the results obtained using the classical Multi Dimensional Voice Program (MDVP), with the results obtained with the novel WPCVox. The aim of this comparison was to find differences and similarities in the parameters extracted with both systems in order to make comparison of measurements and data transfer among both equipments. The study was carried out in two stages: in the first, a wide sample of healthy voices belonging to Spanish-speaking adults from both genders were used to carry out a direct comparison between the results given by MDVP and those obtained with WPCVox. In the second stage, a sample of 200 speakers (53 normal and 173 pathological) taken from a commercially available database of voice disorders were used to demonstrate the usefulness of WPCVox for the acoustic analysis and the characterization of normal and pathological voices. The results conclude that WPCVox provides very reliable measurements which are very similar to those obtained using MDVP, and very similar capabilities to discriminate among normal and pathological voices.

Keywords Normal voice · Pathological voice · Acoustic measurements · WPCVox · MDVP

Introduction

The present work deals with one of the most complete instruments available for the acoustic analysis of voice: the Multi Dimensional Voice Program (MDVP), developed by Kay Elemetrics [1], in comparison with a novel commercial tool that has recently been made available in the market: WPCVox [2]. Both systems are capable of rapidly calculating a large number of parameters to measure different aspects of the vocal function, facilitating the graphical representation of these parameters. Despite of the similarities among both systems, small differences in the estimation of the parameters are expected as those existing between MDVP and the other de facto standard in the field of the acoustic analysis: Dr. Speech [3].

Throughout this paper the normative values are given for the main acoustic parameters given by MDVP and WPCVox with Spanish-speakers. At the same time, the parameters calculated by both systems are compared in order to determine whether the parameters calculated are measuring the same phenomenon or not. Moreover, for the sake of comparison, the discrimination capabilities to separate normal and pathological voices have been analysed with both sets of equipment.

Material and methods

Selection of the corpus speakers

In order to represent the normality criteria, 520 voice samples corresponding to non pathological speakers were

recorded at the “*Príncipe de Asturias*” Hospital in Alcalá de Henares, Madrid. From this dataset 356 speakers were selected matching the inclusion requirements fixed for the study. The distribution by gender is: 226 females and 130 males. The age range goes from 9 to 79 years old, with a mean of 35.5 years. For each speaker, the acoustically balanced utterance “*es hábil un solo día*” was recorded, together with the sustained phonation of vowel ‘a’ for a duration of at least 3 s.

The inclusion criteria fixed to select the speakers to be part of the final corpus were: (a) The subjective feeling that the speaker has no perceived laryngeal pathology; (b) An adequate voice for the age, gender and cultural group of the speaker, and an adequate pitch, timbre, volume and flexibility of diction; (c) Non smoker; (d) The speaker has never undergone any surgical operation related with any laryngeal pathology; (e) The speaker has not required endotracheal intubations in the last year; (f) A perceptual judgement according to the GRBAS scale with a score below 3.

The corpus was complemented with pathological voices from a wide range of disorders. The same material was recorded for those patients that visited the ENT service of the mentioned Hospital. The pathological dataset contains 203 samples with a wide range of pathologies. Such pathological dataset was not used in this study but will be helpful for future research.

For the second part of the study presented in this paper, we have used the Voice Disorders Database distributed by Kay Elemetrics. Two hundred and twenty six samples were selected according to the criteria explained (53 normal and 173 pathological). This database was used to test the usefulness of WPCVox for acoustic analysis and discrimination among normal and pathological voices. The experiment was carried out with the database distributed by Kay Elemetrics because it is easily available and it would allow the reproducibility of the results.

The explanation to the fact that the databases used have been recorded with the MDVP is that in order to compare the algorithms implemented in both equipments, the speech files had to be recorded with the same system to avoid differences that could appear due to the A/D converter or sampling rates.

Description of the exploration techniques

The exploration techniques followed to obtain the normal voices consist of several steps, first of all, the vocal anamnesis is carried out, paying special attention to those factors that do not match the inclusion criteria enumerated before; after the anamnesis, a stroboscopy with a telelaryngoscope was usually carried out to dismiss any laryngeal

pathology;¹ then the microphonic joint to the electroglottographic signal was recorded. A condenser microphone was used, placed on a stand at 30 cm from the mouth at a 50° angle. The recordings were performed with the CSL 4300B equipment of Kay Elemetrics, sampling both signals with a frequency of 50 kHz and 16 bits of quantization. All the recordings were made under the same conditions, using macros to ensure an identical configuration of the configuration parameters. All the samples were collected in a soundproof room. The samples correspond to the sustained phonation of the Spanish vowel ‘a’. The first 500 ms. and the last part of the utterance were removed to avoid onset and offset effects. Finally, only 3 s. (the midvowel segment) was used for further analysis.

Regarding the second stage of the work, 173 pathological and 53 normal speakers were selected according to the criteria fixed. After edition and selection of the stationary part in the middle of the sustained phonation, the files were stored in the database by the developers with a sampling frequency of 50 kHz. The duration of the pathological records is around 0.8 s. long, whereas the normal voices are 3 s. long. The samples correspond to the sustained phonation of the vowel ‘ah’.

Instrumentation: MDVP and WPCVox equipments

The MDVP is a part of the CSL-4300B equipment. It is probably one of the most standardized systems currently existing in the market. The equipment has its own input/output recording device connected to a PC by means of a PCI hardware interface. The hardware has been developed to simultaneously record four signals with a maximum sampling frequency of 51.2 kHz. With its integrated hardware and software, it is tailored for sound input and measurement in speech processing applications. The system records different types of signals (electroglottographic, microphonic, and inverse filtered signals) providing the possibility of using several software modules that complement the analysis. The MDVP provides up to 24 parameters (frequency and amplitude perturbation, noise, tremor parameters, etc) providing a multidimensional analysis of the voice. The most important parameters calculated by MDVP are summarized in Table 1.

WPCVox is an application developed with the aim of recording and analyzing speech and electroglottographic (EGG) signals. This is a low-cost commercial application that just becomes available to the scientific community. WPCVox is one of the first commercial applications developed

¹ Not all the patients were subjected to an endoscopy; only those presenting a vocal disorder based on a previous psychoacoustic judgement of their voice. In total, 200 patients underwent these exploration techniques.

Table 1 Description of the acoustic measurements performed by WPCVox and MDVP

Type	Parameter	Description	MDVP	WPCVox
Frequency perturbations	fo	Average fundamental frequency for the vocalization (Hz).	×	×
	Jitt abs	Absolute jitter gives an evaluation of the period-to-period variability of the pitch period (ms).	⊗	×
	Jitt rel	Jitter percent represents the relative period-to-period (in short-term) variability of the pitch (%).	⊗	×
	RAP	Relative average perturbation gives the variability of the pitch period with a smoothing factor of three periods (%).	×	⋅
	PPQ	Pitch period perturbation quotient gives the variability of the pitch period with a smoothing factor of 5 periods (%).	⊗	×
	sPPQ	Smoothed PPQ gives the long-term variability of the pitch period with a user-selected number of periods (usually 55) (%).	×	⋅
Amplitude perturbations	Shimm abs	Shimmer in dB gives the period-to-period variability of the peak-to-peak amplitude within the analysed voice sample (dB).	×	⋅
	Shimm rel	Shimmer percent gives the variability of the peak-to-peak amplitude. It represents the relative period-to-period (in short-term) variability of the peak-to-peak amplitude (%).	⊗	×
	APQ	Amplitude perturbation quotient gives the variability of the peak-to-peak amplitude with a smoothing factor of 11 periods (%).	×	⋅
	sAPQ	Smoothed APQ gives the long-term variability of the peak-to-peak amplitude with a user-selected number of periods (usually 55) (%).	⊗	×
Noise	NHR	Noise-to-harmonics ratio is an average ratio of energy of the in-harmonic components in the range 1,500–4,500 Hz to the harmonic components energy in the range 70–4,500 Hz (%).	×	⋅
	HNR	Harmonics-to-noise ratio is an average ratio of the energy of the harmonics related to the noise energy in the range 70–4,500 Hz (dB).		×
	NNE	Normalized noise energy is a ratio of the energy of the noise present above 1 kHz to the total energy of the signal (dB).		×
	VTI	Voice turbulence index is an average ratio of the spectral in-harmonic high-frequency energy in the range 2,800–5,800 Hz to the spectral harmonic energy in the range 70–4,500 Hz (%).	⊗	⊗

in Europe for the objective analysis of voice disorders. The recording is carried out using an external standard sound card (SoundBlaster Live 24 bit External) connected to a USB port that gives a signal to noise ratio (SNR) over 100 dB with 24 bits of resolution and sampling frequencies above 48 kHz. This software tool permits the synchronous recording of speech and EGG signals using an active connector specially developed to mix both signals together. The application is able to extract and represent graphically the most useful parameters for evaluating voice quality (Table 1).

Statistics

For each parameter, the samples used to compare the results obtained with both sets of equipment were selected within three times the interquartile range, the samples falling outside this range being regarded as outlayers due to measurement errors. For testing the discrimination capabilities no samples were rejected.

For the sake of comparison, parametric (Pearson correlation coefficient) and nonparametric tests (Kendall and Spearman correlation coefficients) were carried out, considering that there exists a statistical significance with a *P* value below 0.05 (95% confidence²). Moreover, the significance between the differences among the two programs was measured using the *t*-test considering that there exists statistical significance with a *P* value below 0.05 (95% confidence). In the parametric tests, when it was necessary due to the shape of the underlying distribution, the parameters were transformed to match the criteria of gaussianity and homocedasticity using a neperian logarithmic transformation: $\ln(\cdot)$. Normality was tested using the Kolmogorov–Smirnov test considering that there exists statistical significance with a *P* value above 0.05 (95% confidence)-, and Q–Q plots. This transformation allows

² Contrasting is more confident evaluating the statistical significance with 95% than 99%.

the use of parametric tests such as Student's *t*. This solution has been used previously by other authors.

The comparison using correlations among measurements has been used by other authors before. A high correlation is a necessary condition to say that both algorithms are similar. The sufficient condition is found in the fact that the measurements are based on the same grounds and algorithms.

All statistical analyses were performed in SPSS 12.0S for Windows (SPSS Corporation, Chicago, IL, USA, 2003).

Results

The comparative results are grouped into two sections. The first deals with the similarities between the values of the parameters obtained with both systems and the calculation of the normative values. The second is a comparison of the discrimination capabilities of both sets of equipment.

Normative values

This section compares the values obtained for the fundamental frequency, the amplitude and frequency perturbations, and the noise parameters; all of them calculated using both systems. The comparison is done by means of the correlation coefficients among those parameters with a similar definition. Table 5 summarises the normative values for MDVP and WPCVox, giving the mean, mode, the 75 and 90% percentiles. The mode values were included because the histograms do not follow a Gaussian distribution.

Fundamental frequency (fo)

The mean fundamental frequency (fo) obtained with WPCVox is around 120 Hz for males and 200 Hz for females, both groups following a Gaussian distribution. As expected, significant variations have been found in both groups ($P < 0.001$). Table 2 shows the mean (μ), standard deviation (σ), 1 and 99% percentiles of the fo separated by gender for both equipments. In terms of comparison, the results obtained using both equipments are almost identical, presenting a high correlation coefficient ($r = 0.98$, according to Pearson).

Table 2 fo (Hz) measured with MDVP and WPCVox

Gender	Items	μ	σ	Percentile		μ	σ	Percentile	
		MDVP		1%	99%	WPCVox		1%	99%
♀	222	199.98	31.4	194.50	205.47	200.35	30.9	194.97	205.75
♂	126	120.56	24.8	114.77	126.35	119.70	24.6	113.97	125.43

Frequency perturbation: jitter parameters

The measurement with MDVP gives a mean value of jitt abs of 58.85 ms, with a wide range of variation. The jitt abs measured with WPCVox is 67.77 ms, with a higher standard deviation. 90% of the samples with the MDVP are below 119.92 ms, whereas using WPCVox they are less than 135.77 ms (Table 5). The parameter shows a high correlation between both systems (Table 4). Statistically significant differences ($P < 0.001$) have been found between the means using both equipments and separating by gender (Table 3).

Regarding the jitt rel, the mean obtained with MDVP is 0.94%, being significantly higher for the females than for the males (Table 5). Results with WPCVox were significantly different (Table 3), with a mean value of 1.13%. Once again, differences by gender (Table 3) were found (Table 5). The existing correlation between both measurements might be considered high (Table 4).

The mean value with MDVP for RAP is 0.56% and for PPQ is 0.55% (Table 5), below the normative data found in the MDVP manual (0.68% for RAP and 0.84% for PPQ). In the case of WPCVox, the results are slightly higher: 0.69 and 0.66%, respectively (Table 5); being

Table 3 *t*-test (assuming different variances) to determine whether there is or not difference of means separating by gender

Parameter	Equipment	Gender	
		MDVP	WPCVox
Jitt abs (ms)	Yes*	Yes*	Yes ($P = 0.001$)
Jitt rel (%)	Yes*	Yes ($P = 0.040$)	Yes*
RAP (%)	Yes*	Yes ($P = 0.008$)	Yes*
PPQ (%)	Yes*	Yes ($P = 0.047$)	Yes*
sPPQ (%)	Yes*	No ($P = 0.081$)	No ($P = 0.973$)
Shimm abs (dB)	Yes*	Yes ($P = 0.011$)	Yes ($P = 0.004$)
Shimm rel (%)	Yes*	Yes ($P = 0.010$)	Yes ($P = 0.003$)
APQ (%)	Yes*	Yes*	Yes*
sAPQ (%)	No ($P = 0.127$)	Yes*	Yes*
NHR (%)	Yes*	Yes*	No ($P = 0.449$)
VTI (%)	Yes*	No ($p = 0.435$)	Yes ($P = 0.002$)
HNR (dB)	Yes*	Yes*	Yes*
NNE(dB)	—	—	Yes*

HNR with MDVP is calculated as $10 \log(1/\text{NHR})$

* $P < 0.001$

Table 4 Correlation measurements among the parameters using MDVP and WPCVox

Parameter	Kendall	<i>P</i>	Spearman	<i>P</i>	Pearson	<i>P</i>
Jitt abs (ms)	0.807	*	0.893	*	0.852	*
Jitt rel (%)	0.776	*	0.873	*	0.847	*
RAP (%)	0.771	*	0.872	*	0.847	*
PPQ (%)	0.767	*	0.865	*	0.837	*
sPPQ (%)	0.811	*	0.889	*	0.866	*
Shimm abs (dB)	0.771	*	0.927	*	0.928	*
Shimm rel (%)	0.769	*	0.928	*	0.928	*
APQ (%)	0.776	*	0.933	*	0.929	*
sAPQ (%)	0.755	*	0.910	*	0.923	*
NHR (%)	0.243	*	0.333	*	0.308	*
VTI (%)	0.119	0.003	0.161	0.003	0.156	0.004
HNR (dB)	0.331	*	0.470	*	0.473	*

* $P < 0.001$

closer to the previously mentioned values in the MDVP manual. Smoothing with 55 pitch consecutive periods, a mean value of sPPQ = 0.77% is obtained using MDVP, and 0.86% with WPCVox, both below the normative values found in the MDVP manual (1.02%). Statistically significant differences have been found between the means of the three parameters using both equipments, and for RAP and PPQ separating by gender (Table 3). Regarding the concordance of the results given by the MDVP and the WPCVox, high correlation coefficients have been found between the mentioned parameters (Table 4). The spreading is higher with WPCVox.

Amplitude perturbation: shimmer parameters

The amplitude perturbation measurements, known as shimmer parameters, have a relevant clinical significance as their equivalent in frequency. Despite their importance and clear relationship with the degree of dysphonia, the study of shimmer has been less extensive in the literature than the study of jitter.

The normative results, separated by gender, are summarized in Table 5. The mean values of shimm abs and shimm rel obtained with MDVP are 0.28 dB and 3.22%, respectively (Table 5). With WPCVox the results were 0.31 dB and 3.48%, close to the normative values presented in the MDVP manual (0.31 dB and 3.81%). The smoothing with 11 and 55 pitch periods gives mean values of APQ = 2.39% and sAPQ = 4.61% using MDVP, and APQ = 2.57% and sAPQ = 4.72% with WPCVox. The shimmer parameters revealed high correlation coefficients (Table 4). There are statistically significant differences separating by gender for both equipments and there are also differences between the two programs, except for sAPQ (Table 3). The spreading of the values is higher with WPCVox.

Noise measurements

The noise parameters measure the degree of hoarseness caused by air turbulences introduced by the presence of a disorder. There are many noise parameters reported in the literature and WPCVox and MDVP provide different noise measurements (see Table 1). Those measurements that seem to be equivalent in both systems -NHR and VTI- have been directly compared. Also, the HNR calculated with WPCVox has been transformed using $10 \log(1/\text{HNR})$ to be compared with the NHR given by MDVP. This transformation has been carried out previously by other authors.

The normative results, separated by gender, are summarized in Table 5. The normative values obtained for the NHR using both equipments differ enormously (Table 3): MDVP gave mean values around 0.13% (above the normative data reported in the MDVP manual), whereas WPCVox gave 1.05% on average (Table 5). In contrast with WPCVox, there have been found statistically significant differences of means separating the dataset by gender using MDVP (Table 3). Another difference found is related with the shape of the distribution: the values obtained with MDVP follow a Gaussian distribution, whereas the data obtained with WPCVox do not. Focusing on the correlation between the data obtained with WPCVox and MDVP, there is a statistically significant weak relationship between both measurements (Table 4).

Regarding the VTI, the normative values obtained using both equipments also differ greatly (Table 3): MDVP gave mean values around 0.04%, whereas WPCVox gave 0.51% on average (Table 5). In contrast with NHR, statistically significant differences of means have been found with WPCVox separating the dataset by gender, but not with MDVP (Table 3). The correlation between both series of VTI demonstrates that there exists a very weak relationship between the measurements (Table 4).

Table 5 Normative values of the parameters calculated with MDVP and WPCVox

Normal voices	♀ ♂		♀		♂	
	MDVP	WPCVox	MDVP	WPCVox	MDVP	WPCVox
Jitt abs (ms)						
μ	58.85	67.77	51.30	62.55	72.25	77.08
σ	42.61	50.47	37.27	47.84	48.03	53.80
Mode	23.91	24.44	23.91	24.44	39.56	15.63
Percentile 75%	78.82	89.24	68.52	86.29	93.51	98.53
Percentile 90%	119.92	135.77	109.82	126.80	154.55	169.54
Jitt rel (%)						
μ	0.94	1.13	0.99	1.24	0.84	0.96
σ	0.65	0.89	0.70	0.94	0.54	0.78
Mode	0.38	0.45	1.36	0.45	0.31	0.48
Percentile 75%	1.22	1.50	1.39	1.65	1.02	1.08
Percentile 90%	1.89	2.27	2.03	2.36	1.71	2.05
RAP (%)						
μ	0.56	0.69	0.60	0.76	0.49	0.57
σ	0.40	0.55	0.43	0.57	0.33	0.48
Mode	0.25	0.27	0.25	0.29	0.25	0.18
Percentile 75%	0.73	0.93	0.85	1.01	0.62	0.67
Percentile 90%	1.15	1.41	1.21	1.47	1.01	1.23
PPQ (%)						
μ	0.55	0.66	0.58	0.72	0.49	0.56
σ	0.38	0.51	0.41	0.54	0.32	0.44
Mode	0.21	0.31	0.24	0.23	0.26	0.26
Percentile 75%	0.71	0.89	0.78	0.98	0.58	0.63
Percentile 90%	1.12	1.29	1.16	1.36	1.02	1.20
sPPQ (%)						
μ	0.77	0.86	0.75	0.87	0.80	0.86
σ	0.41	0.50	0.40	0.51	0.41	0.48
Mode	0.48	0.36	0.48	0.37	0.88	0.36
Percentile 75%	0.95	1.04	0.96	1.06	0.91	1.01
Percentile 90%	1.28	1.49	1.31	1.48	1.26	1.56
Shimm abs (dB)						
μ	0.28	0.31	0.27	0.29	0.31	0.39
σ	0.14	0.14	0.12	0.12	0.16	0.17
Mode	0.28	0.23	0.28	0.21	0.23	0.23
Percentile 75%	0.33	0.37	0.32	0.36	0.38	0.41
Percentile 90%	0.47	0.51	0.40	0.46	0.52	0.58
Shimm rel (%)						
μ	3.22	3.48	3.06	3.27	3.53	3.86
σ	1.57	1.61	1.41	1.40	1.78	1.87
Mode	1.69	2.25	1.69	2.25	2.85	3.05
Percentile 75%	3.81	4.22	3.70	4.07	4.31	4.77
Percentile 90%	5.45	5.81	4.64	5.22	5.99	6.56
APQ (%)						
μ	2.39	2.57	2.19	2.35	2.75	2.98
σ	1.08	0.15	0.96	1.00	1.19	1.28
Mode	1.48	1.13	1.48	1.31	3.13	1.54
Percentile 75%	2.82	3.08	2.56	2.81	3.20	3.64

Table 5 continued

Normal voices	♀ ♂		♀		♂	
	MDVP	WPCVox	MDVP	WPCVox	MDVP	WPCVox
Percentile 90%	3.94	4.16	3.19	3.68	4.39	4.80
sAPQ (%)						
μ	4.61	4.72	4.32	4.37	5.12	5.34
σ	1.90	2.03	1.67	1.67	2.17	2.44
Mode	2.53	3.10	2.76	4.14	3.61	3.71
Percentile 75%	5.63	5.67	5.36	5.26	6.70	6.54
Percentile 90%	7.23	7.50	6.72	6.74	7.81	8.62
NHR (%)						
μ	0.13	1.05	0.12	1.00	0.13	1.14
σ	0.02	1.92	0.02	1.91	0.02	1.93
Mode	0.13	0.03	0.12	0.03	0.13	0.13
Percentile 75%	0.14	1.10	0.13	1.07	0.15	1.28
Percentile 90%	0.15	2.89	0.14	2.62	0.16	3.27
VTI (%)						
μ	0.04	0.51	0.04	0.40	0.04	0.70
σ	0.01	0.92	0.01	0.76	0.02	1.13
Mode	0.04	0.03	0.04	0.02	0.04	0.07
Percentile 75%	0.05	0.56	0.05	0.42	0.05	0.75
Percentile 90%	0.06	1.40	0.06	1.11	0.06	1.93
HNR (dB)						
μ	29.07	25.62	29.23	26.29	28.79	24.40
σ	0.77	3.53	0.75	3.35	0.75	3.53
Mode	29.36	26.01	29.36	26.01	28.79	21.95
Percentile 75%	29.51	28.07	29.59	28.66	29.16	26.59
Percentile 90%	30.19	30.23	30.31	30.68	29.78	29.58
NNE (dB)						
μ	–	–15.28	–	–16.44	–	–13.21
σ	–	0.92	–	4.73	–	18.17
Mode	–	–21.48	–	–26.87	–	–9.08
Percentile 75%	–	–11.97	–	–12.89	–	–9.96
Percentile 90%	–	–8.69	–	–9.78	–	–7.89

With respect to HNR, the normative values obtained using both equipments also differ from the statistical point of view (Table 3): MDVP gave mean values around 29.07 dB, whereas WPCVox gave 25.62 dB on average (Table 5). Moreover, statistically significant differences of means have been found both with WPCVox and MDVP separating the dataset by gender. The study of the correlation between both series of HNR demonstrates that there exists only a weak relationship between both measurements (Table 4).

Regarding the measurements of NNE made with WPCVox, no comparisons can be carried out because MDVP do not provide such a parameter. Its normative values are summarized in (Table 5). Once again, the spreading of the values is clearly higher with WPCVox.

Discriminative capabilities

The discriminative capability of the acoustic parameters given by WPCVox and MDVP has been tested and compared. This test has been carried out with the Voice Disorders Database distributed by Kay Elemetrics. The parameters listed in Table 6 were computed for every register, and the results analysed in two ways: comparing the probability distributions of normal and pathological voices for each parameter and computing the receiver operating characteristics (ROC) curves (Fig. 1). The ROC is analyzed calculating the area under the curve (AUC) and the standard error (SE) of the area as suggested.

Table 6 shows the classification rate of the parameters obtained from the histograms at the threshold value that

Table 6 Classification rates, AUC and SE for the set of parameters calculated with MDVP and WPCVox

Type	Parameter	Classification rate (%)		AUC		SE	
		MDVP	WPCVox	MDVP	WPCVox	MDVP	WPCVox
Frequency perturbation	Jitt abs (ms)	83.41	82.24	0.89	0.86	0.021	0.025
	Jitt rel (%)	75.11	73.83	0.84	0.80	0.027	0.031
	RAP (%)	75.11	74.30	0.84	0.79	0.027	0.032
	PPQ (%)	74.19	73.40	0.84	0.79	0.026	0.032
	sPPQ (%)	76.49	75.70	0.84	0.82	0.027	0.030
Amplitude perturbation	Shimm abs (dB)	83.41	82.71	0.92	0.92	0.018	0.019
	Shimm rel (%)	83.41	83.64	0.92	0.92	0.018	0.019
	APQ (%)	84.79	85.04	0.92	0.92	0.018	0.020
	sAPQ (%)	79.26	80.84	0.90	0.90	0.021	0.021
Noise	NHR (%)	79.26	78.90	0.88	0.84	0.022	0.027
	VTI (%)	63.13	75.70	0.72	0.84	0.037	0.027
	HNR (dB)	79.26	84.57	0.88	0.92	0.022	0.027
	NNE (dB)	–	89.71	–	0.94	–	0.014

separates optimally normal and pathological voices. Moreover, the AUC curve and the SE are presented. In general, the SE shows that the AUC presents a clear overlapping for both systems, so we cannot conclude that there exist evidences that neither WPCVox nor MDVP have better discrimination capabilities. Figure 1 shows the ROC curves for all the parameters extracted with MDVP and WPCVox.

Discussion

The discussion is organized into two sections following the same scheme used to present the results. The first is a discussion of the similarities between the parameters obtained with both systems and the normative values. The second discusses the discrimination capabilities of both equipments.

Normative values

Fundamental frequency

As expected, the variability is higher in the male group than for the females. This fact is clearly reflected in the range and standard deviation calculated (Table 2). The high correlation between the values obtained with both equipments must be emphasized. The mean values and the standard deviation found (Table 2) are very similar to those reported in the state of the art. Moreover, the results are in concordance with those obtained by Karnell.

Frequency perturbation parameters

Among the frequency perturbation parameters, the jitt abs, measured in ms (absolute value), has been one of the parameters most widely used by clinicians, with a long tra-

dition in scientific literature. The results obtained for MDVP and WPCVox regarding the 90% percentile are slightly above the normative values given in the MDVP³ manual (83.2 ms) and match well with the results published by other authors who used Spanish speakers.

The values of jitt rel obtained are higher than those obtained by other researchers: Hollien obtained 0.48% for males with a fo around 102 Hz, and 0.76% for males with fo around 142 Hz. Orlikoff obtained 0.42% for males aged between 26 and 33 years, and González obtained 0.94% for the females and 0.68% for the males subset, using young Spanish speakers with a significant sample size ($n = 148$). Also, the normative values obtained for the jitt rel are higher than those found in other studies done in Spain. However, these studies were carried out with another analysis system (Dr. Speech, by Tiger Electronics).

Fernández, used a dataset of 154 non-smoking adults, finding a value of PPQ = 0.23; the group of smokers in Damborenea yielded PPQ = 0.27. A wide sample of former studies revealed RAP values lower than those obtained in this work. As examples, Preciado's study, using a control group of 64 Spanish teachers, obtained RAP = 0.35; the work of Walton and Orlikoff, obtained RAP = 0.28; Dwire and McCauley, with North American patients, obtained RAP = 0.38 for males and RAP = 0.89 for females, both groups in the age range of 18–25; Takahashi and Koike, using Japanese speakers, achieved a RAP = 0.57 for males and RAP = 0.61 for females.

The significant differences found between genders in the frequency perturbation parameters confirm what is

³ The normative values in the MDVP manual were obtained with English speakers from the sustained phonation of vowel 'ah'

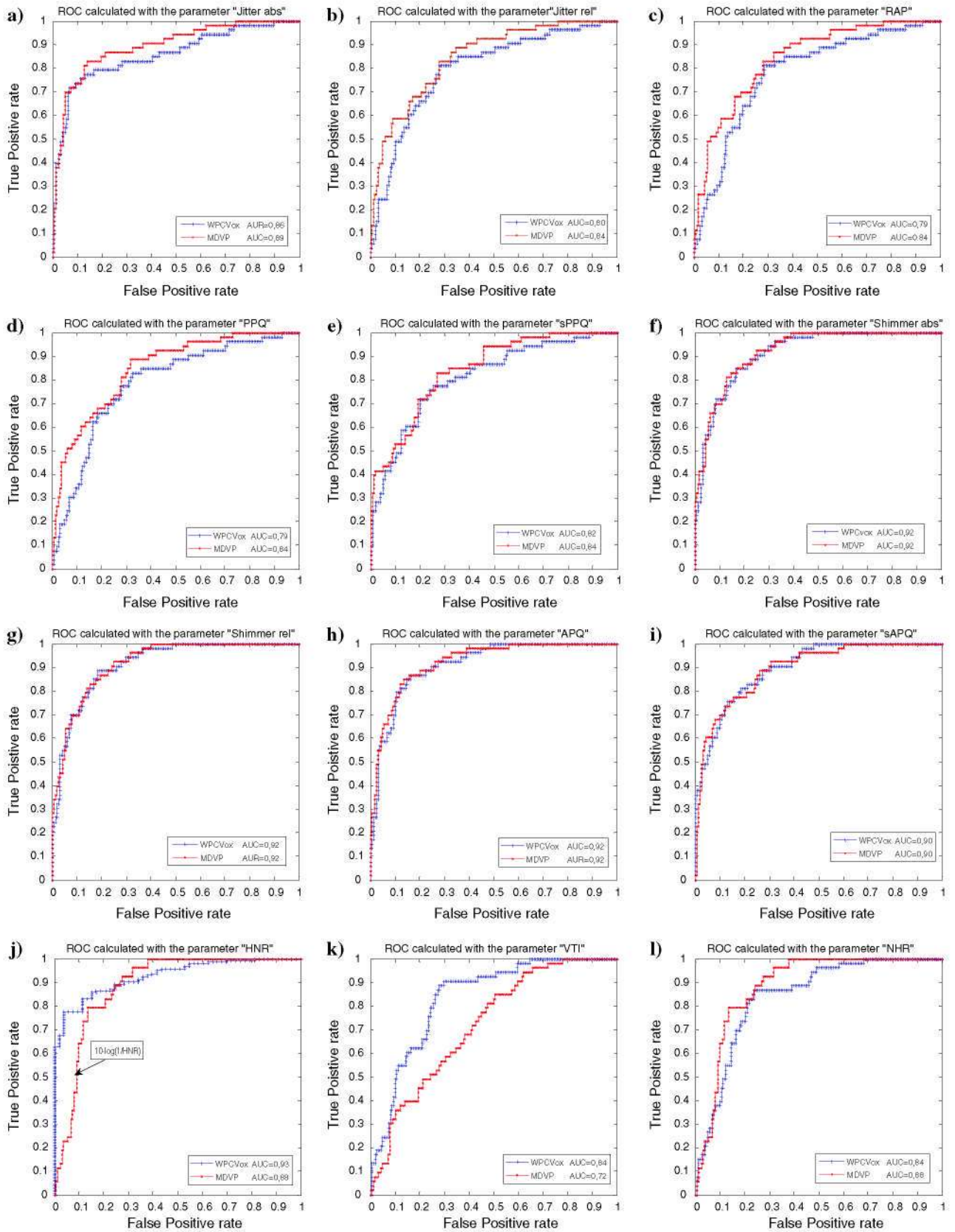


Fig. 1 Receiver operating characteristic curves for the parameters extracted with MDVP and WPCVox; **a**) jitter abs; **b**) jitter rel; **c**) RAP; **d**) PPQ; **e**) sPPQ; **f**) shimmer abs; **g**) shimmer rel; **h**) APQ; **i**) sAPQ; **j**) HNR; **k**) VTI; **l**) NHR

expressed regarding these differences: “the preliminary evidence points out to the possibility that the adult females could have more vocal jitter than males, at least for several vowels. Fortunately, the observed differences between genders are not big enough to provide a diagnostic error”. This fact is also affected by the difference that exists between males and females regarding the f_0 ; despite the normalization carried out according to the f_0 value, there still exist a minor influence.

The fact that WPCVox has just being available in the market makes it difficult to find previous works using this tool in existing literature. There do, however, exist works carried out with other tools: Karnell compared three analysis systems, finding discrepancies in the results obtained for the jitter and shimmer parameters (with a low correlation coefficient $r=0.45$). The only parameter that revealed great correlation was the f_0 . These results revealed that the equipments compared vary more between themselves than do the MDVP and WPCVox. The same study revealed that the correlation between the perturbation parameters obtained were partially improved by rejecting the sample outlayers. In any case, the correlation index was lower than the correlation given between MDVP and WPCVox.

Despite of the differences of the mean values, the high correlation coefficients obtained by comparing the frequency perturbation results between MDVP and WPCVox, and the fact that the parameters are calculated using the same algorithms in both systems, revealed a great concordance. The differences have been laid down to the fact that the algorithms are based on the same grounds but small differences must be expected in their implementation.

Amplitude perturbation parameters

Horii used for his study the ‘a’ vowel uttered by 31 males and 20 females, obtaining shimmer abs equal to 0.47 dB and 0.33 dB, respectively, with no significant differences among genders. In our study, 0.31 dB were obtained for the males group with WPCVox and 0.39 dB for the females, with statistically significant differences between the genders. The sample size must be responsible for the differences. Orlikoff and Kahane reported an average for normal male voices around 0.31 dB.

Comparing with other studies that used a corpus of Spanish speakers, in González, the 90% of the shimmer abs values were below 0.53 dB and shimmer rel values below 5.98%. These ranges are smaller when analyzing our dataset with both programs (Table 5).

The non smoking patients in Fernández achieved an average APQ = 2.02%; the smoking patients in Damborenea, 2.56%; González reported APQ = 2.95% and sAPQ = 5.02%. These values are close to the values

obtained in this study and congruent with the normative data described in the literature.

The high correlation coefficients obtained by comparing the amplitude perturbation results among MDVP and WPCVox, and the fact that the parameters are calculated based on the same algorithms in both systems, revealed a great concordance. Once again, the differences have been laid down to the fact that the algorithms are based on the same grounds but small differences must be expected.

Noise measurements

It has been amply demonstrated that the parameters based on harmonic energy and noise have an extensive relationship with the physiopathological mechanism of dysphonia.

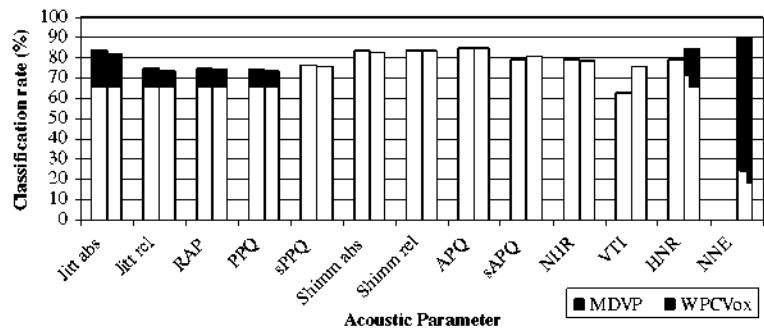
The average NHR measured with the MDVP is 0.13% and with WPCVox 1.05%. A total of 75% of the population studied with MDVP reported values below 0.14 and 90% below 0.15%. These values are below the normative data given by the MDVP manual. The differences of the NHR between females (0.12%) and males (0.13%) using MDVP is statistically significant. The González study reported very similar results and statistical significances. Focusing on the correlation between the data obtained with WPCVox and MDVP, statistically significant similarities have been found to ensure that there is a weak relationship between both measurements.

The VTI measures noise turbulences caused by an incomplete closure of the vocal folds. Hence, most of these noise components are in the high frequency band. The results revealed in our study gave a mean value of 0.04% using MDVP, very similar to the results found in González : 0.05%. Also, González concluded the same about the differences between gender with MDVP. The 75% of the subjects revealed a score below 0.05%; and 90% below 0.06%. Similar results were reported in the MDVP manual. The spreading of the values obtained with WPCVox is 7 times more than the spreading found with MDVP, and the mean values are higher. There is a clear variation between their mean, standard deviation and the shape of the distribution. Moreover, the study of the correlation between both series of VTI demonstrates that there exists a very weak relationship between both measurements.

The comparison between the HNR given by WPCVox and the transformed HNR of MDVP also demonstrated a weak relationship.

The explanation to the weak relationship is that the noise parameters given by both equipments are based on similar principles but measure different aspects of the phenomenon. In contrast to the amplitude and frequency perturbation features, the differences found in the noise parameters stands on the algorithms used by WPCVox and MDVP: whereas MDVP is using the algorithms presented, WPCVox

Fig. 2 Bar plot showing the classification rate of each acoustic parameter measured with MDVP and WPCVox



calculates the noise parameters following the scheme presented. The efficiency of these algorithms has been tested before by their authors, so it will not be treated in this paper. Despite of the differences found, the validity and discrimination capabilities of these algorithms for the clinical diagnosis is clear and will be commented in the next section.

Discrimination capabilities

The results demonstrate the ability to discriminate between normal and pathological voices using the parameters calculated with WPCVox and MDVP. In general, the ability to discriminate between normal and pathological voices may be considered very similar using both equipments. MDVP provides slightly better discrimination capacities using the frequency perturbation parameters, whereas WPCVox seems to work better with noise parameters. The amplitude perturbation parameters showed almost the same discrimination abilities with both equipments.

In spite of the above mentioned differences in measuring the HNR, both equipments showed a clear discrimination capability for this parameter. Also, despite the differences in measuring the VTI, both equipments showed clear discrimination capabilities, with WPCVox being considerably superior for this parameter. Regarding the HNR of WPCVox, it demonstrated better discrimination ability than using the transformed HNR given by MDVP.

The discrimination ability of the parameter NNE is remarkable (89.71%). Such a parameter is not calculated by MDVP.

Figure 2 summarises the discrimination capabilities of the parameters obtained with both systems.

Conclusions

The aim of the study is to prove neither the efficiency nor the accuracy of the algorithms for the detection of voice disorders because they were tested before by other authors. The objective is just to compare the measurements of MDVP and WPCVox to find out whether the values

obtained with both equipments are comparable or not, and to know how large are the differences due to the implementation of the algorithms.

In general terms, the results concluded that WPCVox provides measurements very similar to those obtained with MDVP. The differences found do not demonstrate that WPCVox performs worse or better than MDVP. Moreover, the ROCs demonstrate that the discrimination abilities of both systems are very similar.

Regarding the amplitude and frequency perturbation parameters, the results obtained are also comparable to those found in the literature based on Spanish speakers with other equipments. When we observe the correlation index and compare the acoustic parameters calculated with MDVP and WPCVox, and keeping in mind that both systems are developed following the same algorithms, we can state that there are clear similarities measuring the f_0 and the amplitude and frequency perturbation parameters using both systems.

The worst results in terms of correlation were obtained for the noise parameters: NHR, VTI and HNR. No statistical evidence has been found to indicate that both systems measure the same aspects of the phenomena with these parameters. These differences are explained because the noise parameters of both systems are based on different algorithms. Despite of the differences, it has been demonstrated that the noise measurements of WPCVox have a clear discrimination potential, even better than those of MDVP.

The parameters obtained with WPCVox are reliable for the discrimination between normal and pathological voices, with classification rates around 80% for each parameter alone. WPCVox gives two acoustic parameters (NNE and HNR) not present in MDVP. As reported in the literature, these parameters have revealed great discrimination abilities (greater than the amplitude and frequency perturbation parameters). The NNE's classification rate over 89% must be highlighted.

In conclusion, the matching of most parameters extracted with both equipments (except NHR, VTI and HNR) allows the data transfer possible among them, so those patients evaluated with MDVP could be evaluated

and compared accurately with WPCVox (and vice versa) obtaining very similar results.

Acknowledgments This research was partially carried out under grants: TIC2003-08956-C02-00 from the Ministry of Education of Spain, and AL06-EX-PID-033 from the Universidad Politécnica de Madrid, Spain.

References

- Kay Elemetrics Corporation (1994) MDVP operations manual. Model 4305
- Godino-Llorente JJ, Sáenz-Lechón N, Osma-Ruiz V et al (2006) An integrated tool for the diagnosis of voice disorders. *Med Eng Phys* 28:276–289
- Smits I, Ceuppens P, de Bodt M (2005) A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. *J Voice* 19:187–196
- Hirano M (1981) *Psycho-acoustic evaluation of voice*. Springer, New York
- Kay Elemetrics Corp (1994) *Voice Disorders Database*, version 1.03. Lincoln Park, NJ
- Parsa V, Jamieson D (2000) Identification of pathological voices using glottal noise measures. *J Speech Lang Hear Res* 43:469–485
- Childers D (2000) *Speech processing and synthesis toolboxes*. Wiley, New York
- Baken RJ, Orlikoff R (2000) *Clinical measurement of speech and voice*. 2nd edn. Singular Publishing Group, San Diego, CA, USA
- Feijoo S, Hernández-Espinosa C (1990) Short-term stability measures for the evaluation of vocal quality. *J Speech Hear Res* 33:324–334
- Deliyski D (1993) Acoustic model and evaluation of pathological voice production. In: *Proceedings of Eurospeech'93*. Berlin, Germany 3:1969–1972
- Yumoto E, Gould W, Baer T (1982) Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am* 71:1544–1550
- de Krom G (1993) A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J Speech Hear Res* 36:254–266
- Kasuya H, Ogawa S, Mashima K, Ebihara S (1986) Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J Acoust Soc Am* 80:1329–1334
- Weiss NA (2000) *Introductory statistics*. 6th edn. Addison Wesley, Reading, USA
- Michaelis D, Fröhlich M, Strube HW (1998) Selection and combination of acoustic features for the description of pathologic voices. *J Acoust Soc Am* 103:1628–1639
- Fröhlich M, Michaelis D, Strube HW, Kruse E (2000) Acoustic voice analysis by means of the hoarseness diagram. *J Speech Lang Hear Res* 43:706–720
- Hanley J, McNeil B (1983) A method of comparing the areas under receiver operating characteristics curves derived from the same cases. *Radiology* 148:839–843
- Fernández R, Damborenea J et al (1999) Acoustic analysis of the normal voice in nonsmoking adults. *Acta Otorrinolaringol Española* 50:134–141
- Damborenea J, Fernández R et al (1999) The effect of tobacco consumption on acoustic voice analysis. *Acta Otorrinolaringol Española* 50:448–452
- Karnell MP, Hall KD, et al (1995) Comparison of fundamental frequency and perturbation measurements among three analysis systems. *J Voice* 9:383–393
- Baker K, Ramig L, Jones S, Freed CR (1997) Preliminary voice and speech analysis following fetal dopamine transplants in 5 individuals with Parkinson disease. *J Speech Lang Hear Res* 40:615–626
- McAllister A, Sundberg J, Hibi S (1998) Acoustic measurements and perceptual evaluation of hoarseness in children's voices. *Logopedics Phoniatrics Vocol* 23:27–38
- Wolfe V, Cornell R, Palmer C (1991) Acoustic correlates of pathologic voice types. *J Speech Hear Res* 34(3):509–516
- Wolfe V, Martin D (1997) Acoustic correlates of dysphonia: type and severity. *J Commun Disord* 30(5):403–416
- Preciado JA, Fernández S (1998) Digital analysis of the acoustic signal in vocal pathology diagnosis. Sensitivity and specificity of shimmer and jitter measurements. *Acta Otorrinolaringol Española* 49:475–481
- Hollien H, Thompson CL, Cannon B (1973) Speech intelligibility as a function of ambient pressure and Heo 2 atmosphere. *Aerosp Med* 44:249–253
- Orlikoff R, Baken RJ (1990) Considerations on the relationship between the fundamental frequency of phonation and vocal jitter. *Folia Phoniatrica* 42:31–40
- González T, Cervera T, Miralles J (2002) Análisis acústico de la voz: fiabilidad de un conjunto de parámetros multidimensionales. *Acta Otorrinolaringol Española* 53:256–268
- Orlikoff R, Walton JH (1994) Speaker race identification from acoustic cues in the vocal signal. *J Speech Hear Res* 37:738–745
- Dwire A, McCauley R (1995) Repeated measures of vocal fundamental frequency perturbation obtained using the Visi-pitch. *J Voice* 9(2):156–162
- Takahashi H, Koike Y (1976) Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Otolaryngol* 338:1–24
- Karnell MP (1991) Laryngeal perturbation analysis: minimum length of analysis window. *J Speech Hear Res* 34:544–548
- Horii Y (1979) Fundamental frequency perturbation observed in sustained phonation. *J Speech Hear Res* 22:5–19
- Orlikoff R, Kahane JC (1991) Influence of mean sound pressure level on jitter and shimmer measures. *J Voice* 5:113–119